Chemistry & Biology, Vol. 9, 143, February, 2002, ©2002 Elsevier Science Ltd. All rights reserved. PII S1074-5521(02)00107-2

Preview

Custom Codons Come in Threes, Fours, and Fives

In a laboratory coselection of reporter messages containing a single randomized essential two-, four-, five-, or six-base "codon" with suppressor $tRNA^{Ser}$ libraries whose members possessed randomized anticodon loops of varying sizes, only four- and five-base "codon-anticodon" interactions survived. These suppressor tRNAs accomplish +1 and −1 frameshift suppression, suggesting biological significance. They also display some properties common to serine tRNAs; such properties include a modest excess of Ser anticodons that might assist tRNA charging.

The paper by Anderson et al. on pages 237–244 of this issue demonstrates that *E. coli*'s ribosome can accommodate variably sized codon/anticodon interactions from the usual three up to five base pairs and that these may represent upper and lower bounds. These authors had previously selected tRNAs with anticodon loops that recognize four-base codons to suppress site-specific +1 frameshifts [1] and four-base codon suppressors also exist in nature [2], but five is the newest addition. Since the newly selected tRNAs that recognize five-base codons effectively suppress −1 frameshifts, one wonders whether any rare tRNA-like molecules with unassigned functions might occasionally do this in nature, either as a sophisticated form of error correction or as a conduit to translation in multiple reading frames.

Curiously, Crick and colleagues [3] had suggested much earlier the notion of five-base codon/anticodon interactions. They conjectured that such longer interactions might have been strong enough to support a very early translation system, before ribosomes evolved and before translation was refined. Their proposed *quintuplet* code actually involved two stacking bases flanking a triplet that specified the amino acid. This constrained the sets of possible (overlapping) quintuplets that could participate in such a code and thus restricted the set of amino acids that could have been templated by such a scheme, rather than enhancing it, as is the goal of Anderson and colleagues.

But evolution, true bioengineering, also likely proceeded by code expansion, and several chemists, taking on the challenge, have recently made great strides in using either artificial evolution or design to introduce new ("21^{st}") amino acids into templated proteins [4, 5]. The efficiency of incorporating standard or nonstandard amino acids at three- (often stop codons or under-used codons), four-, or five-base codons varies. The levels for templated serine incorporation by the quintuplets in Table 2 of Anderson et al. (2002) are not higher than 12%, whereas the record for site-specific incorporation of a new amino acid in *E. coli* at an amber stop codon is 67% [4] and the levels in more recent experiments are lower (see [5]). This implies a need for more tinkering if five-base or other custom codons will be an efficient method for introducing nonstandard amino acids.

The specialized context of the modern ribosome stabilizes recognition of three-base codons, but longer ones would have had an advantage in a fierce prebiotic era. Anderson et al.'s experiments raise the tantalizing questions of whether longer codons may have ever had their day (or even several millennia) in a much earlier stage of code evolution and why the number three won. The usual argument that triplets provide sufficient information (but not too many letters) to encode a broad range of amino acids with some degeneracy (permitting evolvability of the sequences) implies either optimization of the current system by natural selection or that three was a very fortunate number that may have resulted from other factors such as stereochemistry. Although 2 nt codons would permit templating of a broad set of amino acids and proteins made of restricted amino acid sets still have complex protein folds [6], it is perhaps surprising that no 2 nt codons arose in Anderson et al.'s experiment. The structures of standard and extended codon-anticodon pairs in the ribosome may answer some of these questions, at least regarding how the modern translation apparatus interprets codon-anticodon pairs of so many sizes.

Laura F. Landweber
Ecology and Evolutionary Biology
Princeton University
Princeton, New Jersey 08544

Selected Reading

1. Magliery, T.J., Anderson, J.C., and Schultz, P.G. (2001). Expanding the genetic code: selection of efficient suppressors of four-base codons and identification of "shifty" four-base codons with a library approach in *Escherichia coli*. J. Mol. Biol. *307*, 755–769.
2. Atkins, J.F., Weiss, R.B., Thompson, S., and Gesteland, R.F. (1991). Towards a genetic dissection of the basis of triplet decoding, and its natural subversion: programmed reading frame shifts and hops. Annu. Rev. Genet. *25*, 201–228.
3. Crick, F.H.C., Brenner, S., Klug, A., and Pieczenik, G. (1976). A speculation on the origin of protein synthesis. Orig. Life *7*, 389–397.
4. Furter, R. (1998). Expansion of the genetic code: site-directed *p*-fluoro-phenylalanine incorporation in *Escherichia coli*. Protein Sci. *7*, 419–426.
5. Böck, A. (2001). Invading the genetic code. Science *292*, 453–454.
6. Riddle, D.S., Santiago, J.V., Bray-Hall, S.T., Doshi, N., Grantcharova, V.P., Yi, Q., and Baker, D. (1997). Functional rapidly folding proteins from simplified amino acid sequences. Nat. Struct. Biol. *4*, 805–809.